Evolution of the Mouse H-2K Region: A Hot Spot of Mutation Associated With Genes Transcribed in Embryos and/or Germ Cells

Young Il Yeom, Kuniya Abe[1] and Karen Artzt

Department of Zoology, The University of Texas at Austin, Austin, Texas 78712-1064

Manuscript received August 15, 1991
Accepted for publication November 16, 1991

ABSTRACT

Active gene transcription is known to promote genetic change in neighboring DNA. We reasoned that the change would be readily heritable if transcription was occurring in germ cells or early embryonic cells before the germ cells are set aside. The H-2K region of the major histocompatibility complex (MHC) provides a good vehicle for testing this hypothesis because it is replete with such genes. We have compared the amount of polymorphism in 240 kb of DNA contiguous with *H-2K* and 150 kb of DNA flanking a homologous duplicated region in *t*-haplotypes and inbred strains. Using 90 probes and three restriction enzymes, we find a staggering difference in the amount of polymorphism in the H-2K region *vs.* the duplicated region (26% *vs.* 0%) of *t*-haplotypes. The disparity in the rate of divergence between the two regions indicates that the spatial distribution of genes and their expression pattern might be important factors in sequence evolution. Since *t*-haplotypes normally show extremely limited variability among themselves due to their recent divergence from a single ancestor, these results imply that the mutation rate in the H-2K region is unusually high. This is in apparent contradiction to the current view that the MHC loci have evolved at the same rate as other loci. The implications for the evolution of the *H-2K* gene are discussed.

THE *t*-complex occupies a 15-cM region at the proximal end of chromosome *17* of the mouse, covering the loci of *T* (*Brachyury*), *qk*, *tf*, and the major histocompatibility complex (MHC) (reviewed in BENNETT 1975; SILVER 1985; KLEIN 1986; SHIN 1989). In about 20% of feral mice, the *t*-complex exists in a variant form called *t*-haplotypes. Various features are shared by all *t*-haplotypes and distinguish them as a whole from wild-type chromosomes. In particular, males heterozygous for a complete *t*-haplotype and a wild-type chromosome typically show distorted transmission of the *t*-bearing chromosome which is preferentially transmitted to up to 99% of the offspring. This phenomenon, called transmission ratio distortion (TRD), has been the major driving force in maintaining the *t*-haplotypes in wild populations.

Despite their relatively high frequency in feral populations, *t*-haplotypes are quite secluded from genetic recombination. Classical and molecular genetic experiments have shown that the length of the *t*-complex is defined by four non-overlapping inversions relative to the wild type (ARTZT, MCCORMICK and BENNETT 1982; SHIN *et al.* 1983; HERRMANN *et al.* 1986; HAMMER, SCHIMENTI and SILVER 1989). It is generally believed that these inversions are responsible for suppressing recombination in *t*-haplotype heterozygotes to a level as low as 0.1% of normal. In contrast, recombination between two *t*-haplotypes can occur at normal frequency throughout the *t*-complex since all *t*-haplotypes share the same inversions (SILVER and ARTZT 1981). However, in natural populations, the latter type of recombination is severely limited for several reasons. First, most *t*-haplotypes carry one or more recessive embryonic lethal mutations precluding the survival of homozygotes. Although there are at least 16 different complementing recessive lethals (KLEIN, SIPOS and FIGUEROA 1984), these are rarely found in the same population. Second, even if different lethals existed in the same population, male compound heterozygotes (t^x/t^y) are sterile, and the number of all compound males and females would be reduced further because different *t*-lethals do not complement each other fully. Thus, the opportunity to observe recombination in *t* chromatin is severely restricted. *t* Chromosomes show very limited variability which is generally believed to be due to their common descent from a single ancestor 1–2 Myr ago (SILVER *et al.* 1987). Thus, the entire *t*-complex has evolved virtually as a single genetic unit in the *t*-haplotypes, relatively isolated from the pool of chromosome *17* genes in the mouse, and also partially isolated from one another. Because of this genetic isolation, the *t*-complex provides fruitful material for studying the evolution of the mammalian genome.

It has been demonstrated that active transcription of a gene can increase the mutation rate in the nearby sequence of eukaryotic DNA (BOHR, PHILLIPS and HANAWALT 1987; a review). This is presumably be-

Y.I.Y. dedicates this paper to the memory of his mentor, DOROTHEA BENNETT (1929–1990).

[1] Present address: Institute for Medical Genetics, Kumamoto University Medical School, 4-24-1 Kuhonji, Kumamoto 862, Japan.

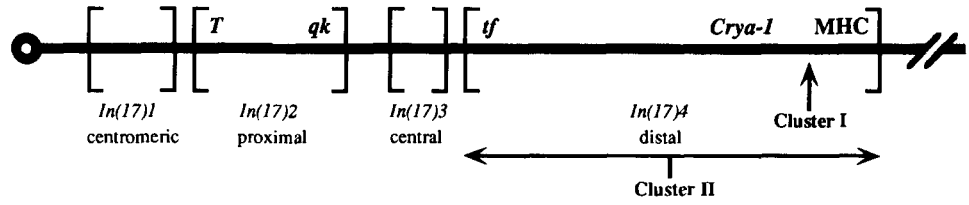


FIGURE 1.—The map position of clusters I and II on chromosome 17. The map was drawn in the wild-type orientation. Markers such as *T (Brachyury)*, *quake (qk)*, *tufted (tf)*, *α-crystallin (Crya-1)* and the MHC are indicated. Brackets enclose the *t*-complex inversions. Cluster I is physically linked to the H-2K region, whereas cluster II maps within the distal inversion. Cluster II was mapped using a series of partial *t*-haplotypes and a probe polymorphic between *t* and wild type (YEOM *et al.* 1992).

cause transcriptional activity renders the structure of local chromatin more susceptible to insertion of mobile genetic elements or accessible to DNA modifying enzymes and mutagens. We would expand this idea to emphasize the importance of the tissue specificity of transcription with respect to the heritibility of coincident genetic changes. It is proposed that mutations that occur within active genes or in their flanking regions are especially likely to be inherited if the genes are expressed in germ cells or in the early embryo before the germ cell progenitors are set aside. This notion would predict that regions of chromosome which are rich in genes transcribed in germ cells or very early embryos could appear as "hot spots" of genetic change.

We studied the patterns of DNA sequence variation in the H-2K region and in a partially homologous, duplicated region in a panel of *t*-haplotypes and inbred strains. The approach was to analyze restriction fragment length polymorphisms (RFLPs) with a battery of genomic DNA probes cloned from a chromosomal walk. These experiments enabled us to test the above hypothesis because the H-2K region is densely populated with genes transcribed in germ cells and/or in the early embryo, whereas the homologous duplicated region is relatively inactive in transcription (ABE *et al.* 1988; YEOM *et al.* 1992). We show that in both *t*-haplotypes and wild-type homologs the 240 kb of DNA flanking the H-2K region is highly polymorphic, whereas the duplicated region has a low variability.

MATERIALS AND METHODS

Mice: Mice bearing *t*-haplotypes are congenic on the C3H/DiSn (C3H) background, and are bred in our mouse colony. t^6, t^{12}, t^{w5} and t^{w12} are reviewed in BENNETT (1975). t^{w5g} is a viable revertant isolated as a normal-tailed offspring from a tailless C3H.T/t^{w5} × C3H.T/t^{w5} mating; it appears to be identical to t^{w5} on all criteria except that it no longer carries the embryonic lethality. Since homozygotes for t^{w5g} are fully viable, it is a convenient source of *t*-homozygous DNA (K. ARTZT and D. BENNETT, unpublished). t^{w120} is described by ARTZT *et al.* (1985). t^{w130} is a member of the t^{w5} complementation group; it was found in Peru and was generously provided by MARGARET WALLACE. Inbred strains were purchased from The Jackson Laboratory (Bar Harbor, Maine) except for AKR which was a gift of PAUL D. GOTTLIEB.

Probes: Cosmid fragments used as probes in this study are described in UEHARA *et al.* (1987) (for probes 1–38) and YEOM *et al.* (1992) (for probes 39–90).

Restriction mapping of cosmids: Cosmid clones obtained from the chromosomal walk were restriction mapped by the terminase method (RACKWITZ *et al.* 1984, 1985); the detailed maps will be described elsewhere (YEOM *et al.* 1992).

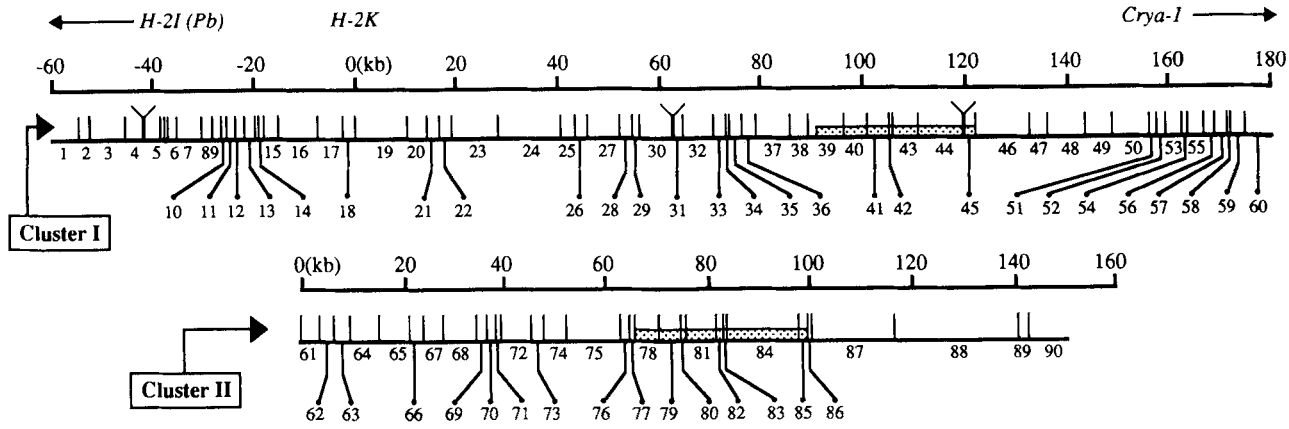
Southern analysis: Restriction-digested DNA was size-fractionated on a 0.7% agarose gel in Tris-phosphate buffer and Southern-transferred to a nylon membrane (Hybond N, Amersham). Blots were hybridized at 42° in a mixture containing 50% formamide/50 mM sodium phosphate (pH 6.5)/5 × SSC/5 × Denhardt's solution/0.1% sodium dodecyl sulfate (SDS)/1 mM EDTA/7.5% dextran sulfate and heat-denatured salmon sperm DNA (100 μg/ml). Probes were ^{32}P-labeled by either nick translation (RIGBY *et al.* 1977) or random hexamer priming (FEINBERG and VOGELSTEIN 1983) to a specific activity of 1–10 × 10^8 cpm/μg DNA. Genomic probes were pretreated prior to hybridization, with 1.5 × 10^4 times excess (w/w) amount of unlabeled mouse DNA at 42° in order to block highly repetitive sequences. Blots were washed in 2 × SSC/0.1% SDS for 50 min at room temperature and then, in 0.1 × SSC/0.1% SDS for 40 min at 65°.

RESULTS

Characterization of the two clusters studied: Two separate clusters of DNA were cloned from the *t*-complex by chromosome walking in the t^{w5} haplotype. These two DNA regions are designated clusters I and II, respectively, and each is defined by a distinctive set of overlapping cosmid clones. Details of the cloning are reported elsewhere (YEOM *et al.* 1992). Briefly, cluster I consists of 240 kb mapping to the H-2K region, whereas cluster II comprises 150 kb located elsewhere in the distal inversion of the *t*-complex. Their map position relative to known genetic markers is shown in Figure 1.

The relationship of clusters I and II can be characterized in three ways. First, both of them map within the distal inversion of the *t*-complex. Second, about 35 kb of DNA in each cluster is a homologous duplicated copy of the other one (see stippled areas in Figure 2). And third, the two clusters have a different coding content in that cluster I contains at least 12 transcribed sequences, while cluster II has only three to four genes. These characteristics, in conjunction with the genetic isolation of *t* chromosomes, provide a unique opportunity to study the evolution of a segment of mammalian genome.

The H-2K region is highly polymorphic, in con-



Probe #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Size (kb)	13.5	2.2	7.1	3.7	3.1	1.7	4.9	2.1	1.7	1.0	1.9	1.7	2.1	1.0	2.8	7.6	5.2	2.4	10.3	3.8	2.5	2.4	8.9	12.4	2.8	2.6	6.5	2.6	1.4	6.6	1.7	6.2	2.3	0.8	2.6	2.9	7.1	3.6	6.8	4.6	4.4	0.8	5.0	9.1	2.2
RFLP	-	+	-	-	-	-	-	-	+	+	-	-	nd	+	+	+	nd	nd	+	+	+	nd	+	+	-	-	+	+	-	-	-	-	+	-	+	-	-	-	-	+	+	-	+	+	+
t^{w5}	13.5	2.2	7.1	3.7	3.1	1.7	4.9	2.1	1.7	1.0	1.9	1.7	2.1	1.0	2.8	7.6	5.2	2.4	10.3	3.8	2.5	2.4	8.9	12.4	2.8	2.6	6.5	2.6	1.4	6.6	1.7	6.2	2.3	0.8	2.6	2.9	7.1	3.6	6.8	4.6	4.4	0.8	5.0	9.1	2.2
t^{12}	"	2.0	"	"	"	"	"	"	2.8	2.8	"	"	nd	"	"	7.1	"	"	10.6	3.0+0.8	3.1	"	8.7	12.8	"	"	6.8	2.8	"	"	"	"	"	"	3.1	"	"	"	"	2.4+2.4	4.45	"	4.6	9.5	"
t^{6}	"	2.2	"	"	"	"	"	"	"	"	"	"	nd	"	2.9	"	nd	nd	"	4.7	nd	nd	nd	12.1	"	"	"	2.6	"	"	"	"	2.4	"	2.6	"	"	"	4.4	"	"	4.8	"	2.3	
t^{w12}	"	"	"	"	"	"	"	"	"	"	"	"	nd	"	"	"	nd	nd	"	"	nd	nd	nd	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"
C3H	"	"	"	"	"	"	"	"	1.7	1.0	"	"	nd	3.9	3.9	7.6	nd	nd	10.3	3.8	nd	nd	nd	12.4	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	2.4+2.4	"	"	4.6	10.0	1.5

Probe #	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
Size (kb)	10.6	3.7	7.6	5.3	7.6	1.4	1.6	3.4	1.2	3.2	2.3	2.7	0.9	3.0	10.3	3.5	2.9	3.3	5.8	6.1	2.7	3.8	6.9	2.3	1.8	1.0	6.1	2.5	4.6	10.8	1.7	1.2	4.6	4.4	1.2	6.0	1.5	0.8	14.0	1.7	0.8	16.6	24.0	2.3	9.0
RFLP	+	+	+	-	+	-	+	-	+	+	+	+	-	-	+	+	+	-	+	-	-	-	-	-	-	-	-	+	+	+	-	nd	-	-	nd	nd	-	-	-	-	-	-	-	-	-
t^{w5}	10.6	3.7	7.6	5.3	7.6	1.4	1.6	3.4	1.2	3.2	2.3	2.7	0.9	3.0	10.3	3.5	2.9	3.3	5.8	6.1	2.7	3.8	6.9	2.3	1.8	1.0	6.1	2.5	4.6	10.8	1.7	1.2	4.6	4.4	1.2	6.0	1.5	0.8	14.0	1.7	0.8	16.6	24.0	2.3	9.0
t^{12}	5.3+4.5	3.75	"	"	8.0	"	"	"	4.5	4.5	5.2	5.2	"	"	8.5	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	nd	"	"	nd	nd	"	"	"	"	"	"	"	"	"	"
t^{6}	5.3+5.3	"	"	"	8.4	"	1.5	"	"	"	"	"	"	"	8.4	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	nd	"	"	nd	nd	"	"	"	"	"	"	"	"	"	"
t^{w12}	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	"	nd	"	"	nd	nd	"	"	"	"	"	"	"	"	"
C3H	"	"	7.4	"	8.0	"	1.7	"	"	"	2.3	2.7	"	"	8.6	2.2	3.8	"	6.1	"	"	"	"	"	"	"	"	2.8	15.0	15.0	"	nd	"	"	nd	nd	"	"	"	"	"	"	"	"	"	"	"	"

FIGURE 2.—Summary of the RFLP pattern of clusters I and II in four *t*-haplotypes and C3H. Cluster I is shown in the *t*-haplotype orientation; the orientation of cluster II is not determined. A *Bam*HI restriction map of the t^{w5} haplotype is shown with fragment sizes indicated under the probe numbers. Stippled regions indicate a pair of duplicated homologous sequences. Probes 20, 40 and 46 each detects two fragments in some haplotypes, whereas each of the probe pairs 9/10, 14/15, 54/55, 56/57 and 74/75 hybridizes with a single band in some chromosomes. Probes 17, 18, 21, 22 and 23 are a part of class I genes, and probes 80 and 81 contain a retroviral sequence; therefore, their restriction fragment sizes were not determined in most of the haplotypes whose genomic DNA was not directly cloned. nd, not determined; +, polymorphic; −, no RFLP.

trast to cluster II: The cosmid DNAs were restriction-digested to derive a collection of cloned genomic DNA fragments for use as probes in Southern analyses of a panel of wild-type and *t*-haplotypes. Figure 3 shows a typical example of the analysis for a wild type (C3H), and five *t*-haplotypes using adjacent 2.3- and 2.7-kb *Bam*HI fragments derived from cluster I of the t^{w5} haplotype (probes 56 and 57 in Figure 2). These two probes detected two *Bam*HI fragments of the same size (2.3 kb and 2.7 kb) in C3H and t^{w5}, but a new 5.2-kb fragment was evident in t^{6}, t^{12} and t^{w12}. Since the sum of 2.3 and 2.7 kb approximates 5.2 kb, it can be inferred that a *Bam*HI site between the 2.3-kb and 2.7-kb fragments in one set of haplotypes might have been lost to give rise to the 5.2-kb fragment in the other set. By combining this type of analysis using several enzymes, we could infer the genomic organization of these other haplotypes without directly cloning and restriction-mapping their genomic DNA.

Figure 2 summarizes the results of Southern analysis in t^{6}, t^{12}, t^{w5}, t^{w12} and C3H with the restriction enzyme *Bam*HI. Two points are clearly evident from these data. First, throughout the analyzed area RFLPs occur frequently between C3H and the four *t*-haplotypes. Nevertheless, the five haplotypes appear to have a similar genomic organization in these regions of the chromosome, without major DNA rearrangements. Second, the occurrence of RFLPs is not random between the two regions analyzed. They occur frequently in the entire 240 kb of cluster I both among *t*-haplotypes, and between *t*-haplotypes and C3H. However, in cluster II the frequency of RFLPs is markedly lower and they are preferentially localized to the left side of the cluster. Furthermore, the limited polymorphism in cluster II is seen only between C3H and *t*-haplotypes, and is not detected among different

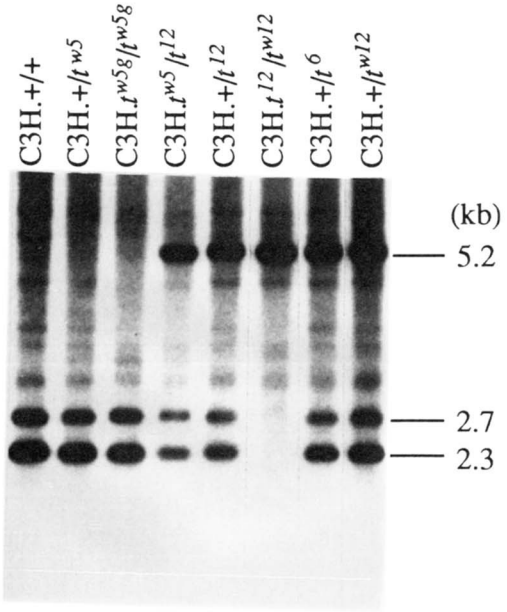


FIGURE 3.—A typical Southern hybridization showing a RFLP in different *t*-haplotypes. The 2.3- and 2.7-kb genomic *Bam*HI fragments derived from the t^{w5} haplotype (probes 56 and 57 in Figure 2) were pooled and hybridized to a blot of *Bam*HI-digested genomic DNA. Repetitive sequences in the probes were blocked by prehybridization with excess unlabeled mouse DNA; however, the background signal is usually higher than with unique-copy probes.

t-haplotypes. These trends were further corroborated by two other enzymes: *Eco*RI and *Kpn*I (data not shown).

The RFLP pattern of clusters I and II in *t*-haplotypes and inbred strains: Table 1 summarizes the results of the RFLP analysis for clusters I and II in the same four *t*-haplotypes using three restriction enzymes: *Bam*HI, *Eco*RI and *Kpn*I. As noted above, the frequency of RFLPs is dramatically different in the two clusters; cluster I shows an average of 26.1% while cluster II has none. The high frequency of polymorphism seen in cluster I of *t*-chromosomes is quite unusual considering that all *t*-haplotypes have descended from a common ancestor, and as a consequence, they rarely show much genetic variation (SHIN *et al.* 1982; SILVER *et al.* 1983; NIZETIC, FIGUEROA and KLEIN 1984; ROEHME *et al.* 1984; SHIN 1989, a review). In this respect, the RFLP pattern in cluster II is more typical of *t*-haplotypes.

Among the four *t*-haplotypes, the t^{w5} chromosome appears to have the most significantly divergent DNA sequences in cluster I (Table 1); this is consistent with the fact that the *H-2* haplotype of t^{w5} is rather unique compared to those of many other *t*-haplotypes (ARTZT *et al.* 1985). In contrast, t^{6} and t^{w12} chromosomes are identical to each other, showing no RFLPs among 200 DNA fragments tested; the identity of the t^{6} and t^{w12} chromosomes in cluster I acts as an internal control because they are known to share the same *H-2* haplotype.

We were interested in whether this RFLP distribution pattern is unique to *t*-haplotypes or is a general phenomenon also seen in wild-type chromosomes. To address this question a comparable analysis was performed on seven inbred strains with five different *H-2* haplotypes (Table 2). (Two strains with redundant *H-2* haplotypes were included as controls.) The analysis of these inbred strains reveals essentially the same phenomenon as observed for *t*-haplotypes. Cluster I sequences show a high degree of polymorphism, whereas cluster II sequences are much less polymorphic (32.1% *vs.* 2.2% on average). These data suggest that the high frequency of RFLP observed in cluster I is not due to the unique evolutionary history of *t* chromosomes, but might be caused by characteristics intrinsically associated with the H-2K region.

Comparison of the RFLP pattern between wild types and *t*-haplotypes: *t*-Haplotypes and their wild-type homologs have taken quite distinctive evolutionary pathways ever since they diverged some 2 million years (myr) ago. Quantitative evidence for this can be found in the data in Table 3 where the RFLP frequencies for the four *t*-haplotypes and seven inbred strains were compared among themselves and to each other. In general, different *t*-haplotypes look more related to one another as a group than to the wild-type strains, since the percent of RFLP between *t*-haplotypes and wild types is always greater than that between different *t*-haplotypes or between different wild-type strains. This trend is especially evident in the RFLP distribution pattern of cluster II, but might be obscured considerably in cluster I due to its highly polymorphic nature. Despite the apparent dissimilarities between the DNA sequences of the *t*-haplotypes and wild types, the overall chromosomal organization as defined by the probe order appears the same in both groups of mice.

It is worth noting that for a specific region, the percentage of RFLP between *t*-haplotypes is significantly lower than between wild types (26.1% *vs.* 32.1% for cluster I, 0.0% *vs.* 2.2% for cluster II; Table 3). Consistent with previous reports, this observation may represent the limited genetic variability of *t* chromosomes due to the unique constraints inflicted on them during their evolution.

Comparison of the RFLP pattern in a pair of duplicated homologous regions: Clusters I and II share a region of 35 kb of duplicated DNA indicated by the stippled areas on the maps of the two clusters in Figure 2. These homologous areas are compared in Figure 4. Although their restriction maps are completely different, the relative order of the corresponding DNA fragments appears to remain unchanged. However, the two homologous segments have been subject to different evolutionary processes after the initial duplication event, since the cluster II copy suffered an insertion of a retroviral sequence (denoted

TABLE 1

RFLP frequencies between four *t*-haplotypes

	Cluster I % RFLP[a]				Cluster II % RFLP[a]			
	t^{w5}	t^{12}	t^{6}	t^{w12}	t^{w5}	t^{12}	t^{6}	t^{w12}
H-2 type:	*w31*	*w28*	*w30*	*w30*	*w31*	*w28*	*w30*	*w30*
t^{w5}		36.6 (134)	32.0 (128)	32.0 (128)	t^{w5}	0.0 (72)	0.0 (72)	0.0 (74)
t^{12}			27.3 (128)	27.3 (128)		t^{12}	0.0 (72)	0.0 (72)
t^{6}				0.0 (128)			t^{6}	0.0 (72)
t^{w12}								t^{w12}

Frequencies were determined using *Bam*HI, *Eco*RI and *Kpn*I.
[a] Percent RFLP is calculated as the No. polymorphic restriction fragments/No. restriction fragments scored × 100. Parentheses enclose the number of restriction fragments scored.

TABLE 2

RFLP frequencies between seven inbred strains

	Cluster I % RFLP[a]							Cluster II % RFLP[a]						
	B6	BALB/c	DBA/2	C3H	DBA/1	SWR	Sm/J	B6	BALB/c	DBA/2	C3H	DBA/1	SWR	Sm/J
H-2 type:	b	d	d	k	q	q	v	b	d	d	k	q	q	v
B6		46.0 (50)	46.0 (50)	28.0 (50)	38.0 (50)	38.0 (50)	46.0 (50)	B6	0.0 (26)	0.0 (26)	3.8 (26)	0.0 (26)	3.8 (26)	0.0 (26)
BALB/c			0.0 (50)	32.0 (50)	32.0 (50)	32.0 (50)	36.0 (50)		BALB/c	0.0 (26)	3.8 (26)	0.0 (26)	3.8 (26)	0.0 (26)
DBA/2				32.0 (50)	32.0 (50)	32.0 (50)	36.0 (50)			DBA/2	3.8 (26)	0.0 (26)	3.8 (26)	0.0 (26)
C3H					26.0 (50)	26.0 (50)	44.0 (50)				C3H	3.8 (26)	7.7 (26)	3.8 (26)
DBA/1						0.0 (50)	36.0 (50)					DBA/1	3.8 (26)	0.0 (26)
SWR							36.0 (50)						SWR	3.8 (26)
Sm/J														Sm/J

Frequencies were determined using *Bam*HI only.
[a] Percent RFLP is calculated as in Table 1. Numbers in the parentheses indicate the numbers of restriction fragments scored.

TABLE 3

Comparison of RFLP frequencies among *t*-haplotypes, among inbred strains, and between the two groups

	% RFLP (No. restriction fragments scored)[a]			
	t vs. t	*t vs. +*	*+ vs. +*	Total
Cluster I	26.1 (774)	36.1 (1672)	32.1 (1050)	32.7 (3496)
Cluster II	0.0 (434)	20.4 (912)	2.2 (546)	10.5 (1892)
Total	16.7 (1208)	30.5 (2584)	21.9 (1596)	

[a] Percent RFLP is calculated as in Table 1.

B-26 envelope *m*ouse retro*v*irus (*Bemv-1*) in Figure 4) (YEOM *et al.* 1992).

The disparity of evolutionary processes weathered by the duplicated sequences is again reflected in their RFLP frequencies; the cluster I copy of the duplicated sequence shows 36.7% RFLP among *t*-haplotypes, while its homologous copy in cluster II shows none (Table 4). (Note that we were unable to test cluster II probes 80 and 81 in the Southern analysis because they contain the retroviral insertion which exists in 40–50 genomic copies.) The RFLP frequency difference between the two homologous regions is also seen among wild types, and between *t*-haplotypes and wild types, making the overall percent RFLP value 42.9% for the cluster I copy and 5.1% for the cluster II copy. It is interesting to note that a pair of homologous DNA sequences located relatively close together, namely within the distal inversion of the *t*-complex, can evolve quite distinctively from each other depending on their genetic environment. Similar observations have been made by several researchers working on the evolution of *H-2* class I and II genes (STEINMETZ *et al.* 1984; GELIEBTER and NATHENSON 1987).

The RFLP pattern of members of the t^{w5} complementation group: The data indicate that cluster I is highly polymorphic and has a strong propensity for being mutated compared to other regions of the genome. To confirm this, the analysis was extended to other members of the t^{w5} complementation group. Figure 5 shows the results of Southern analysis for three members of the t^{w5} complementation group, namely, t^{w5}, t^{w120} and t^{w130} using whole cosmids from cluster I or II as the probe. Although t^{w5} and t^{w120} show an identical restriction fragment pattern for all probes tested regardless if they are from cluster I or II, t^{w130} reveals several RFLPs relative to t^{w5} for cluster I probes (Figure 5, A and B); t^{w130} has the same serological *H-2* haplotype as t^{w12} (K. ARTZT, unpublished data), yet its pattern with cluster I probes is different from t^{w12} (data not shown). The fact that cluster II probes do not detect an RFLP among the

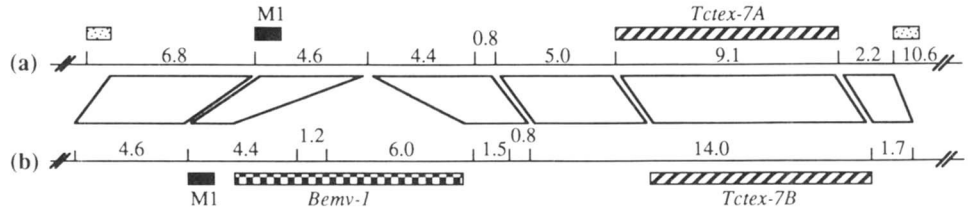


FIGURE 4.—Comparison of the genomic structures for the duplicated homologous region common to cluster I (a) and cluster II (b). The numbers indicate the *Bam*HI fragment sizes in kilobase pairs (kb). Corresponding homologous sequences of the two clusters are depicted as open-boxed areas. Stippled boxes represent a pair of cluster I-specific repeated sequences that define the boundaries of the duplicated area in cluster I; M1 (closed box) is a middle repetitive sequence that is not transcribed. *B*-26 *e*nvelope *m*ouse retro*v*irus (*Bemv-1*) is a 40- to 50-copy insertion which detects multiple messages. It is not known whether this copy is actually transcribed. *Tctex-7* is a four-copy gene family that is transcribed in male germ cells (YEOM *et al.* 1992). The exact molecular boundaries of these sequences are unknown; otherwise the maps are drawn to scale.

TABLE 4

Comparison of RFLP frequencies for the duplicated regions only

	Percent RFLP (No. restriction fragments scored)[a]			
	t vs. t	*t vs.* +	+ *vs.*+	Total
Cluster I	36.7 (120)	48.8 (248)	38.1 (147)	42.9 (515)
Cluster II	0.0 (96)	7.7 (208)	4.8 (126)	5.1 (430)
Total	20.4 (216)	30.0 (456)	22.7 (273)	

[a] Percent RFLP is calculated as in Table 1.

three haplotypes (Figure 5C) further confirms their monomorphism in *t*-haplotypes.

DISCUSSION

We have examined in detail the degree of polymorphism in 240 kb of DNA flanking the *H-2K* gene (cluster I) in *t*-haplotypes and wild-type chromosomes. The H-2K region was compared to 150 kb of a partially homologous duplicated region (cluster II) also located in the distal inversion of the *t*-complex. Southern analysis of the RFLP pattern clearly shows that regardless of the strain, the sequences in cluster I are strikingly polymorphic compared to the sequences in cluster II; 26.1% *vs.* 0% for *t*-haplotypes and 32.1% *vs.* 2.2% for wild type.

The unique evolutionary history of *t*-haplotypes offers a special opportunity to address the issue of genetic change. All *t*-haplotypes are presumed to have been derived from a single ancestral event only 1–2 myr ago (SAGE 1981; NIZETIC, FIGUEROA and KLEIN 1984; FIGUEROA *et al.* 1985; SILVER *et al.* 1987). The product of this event is preserved in an intact form because recombination with wild type is greatly reduced. As a consequence *t*-haplotypes show very restricted genetic variability (reviewed in SHIN 1989). Evidence for this emerges from diverse studies including: (a) serological analyses of the H-2 molecules (STURM, FIGUEROA and KLEIN 1982; NIZETIC, FIGUEROA and KLEIN 1984), (b) Southern blot analyses of the *H-2* class I and II probes (SHIN *et al.* 1982; SILVER 1982; ARTZT *et al.* 1985; FIGUEROA *et al.* 1985) and other *t*-complex probes (GOLUBIC *et al.* 1984; MANN,

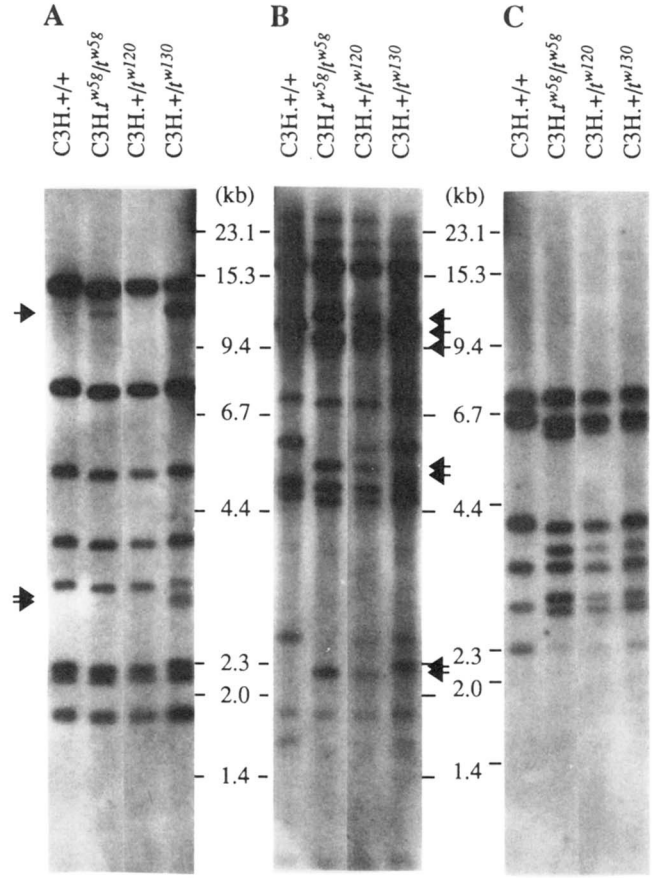


FIGURE 5.—Southern analyses of members of the t^{w5} complementation group. Southerns of *Bam*HI-digested genomic DNA were hybridized with ^{32}P-labeled whole cosmid covering probes 1–9 (A), probes 39–45 (B) and probes 61–68 (C); the positions of the probes are shown in Figure 2. Probe repetitive sequences were blocked as described. Arrows indicate polymorphic bands between the *t*-haplotypes. The faint band located at the first arrow in the t^{w5g} lane of A is believed to be an artifact. DNA size markers are shown in kilobase pairs (kb).

SILVER and ELLIOTT 1986; HA *et al.* 1991; YEOM *et al.* 1992), (c) random DNA probes from the *t*-complex (ROEHME *et al.* 1984), and (d) two-dimensional gel analysis on the testicular cell proteins (SILVER *et al.* 1983). Thus, polymorphic differences between *t*-haplotypes can be ascribed to mutational change in the last 2 myr. What follows is a discussion of resurrecting

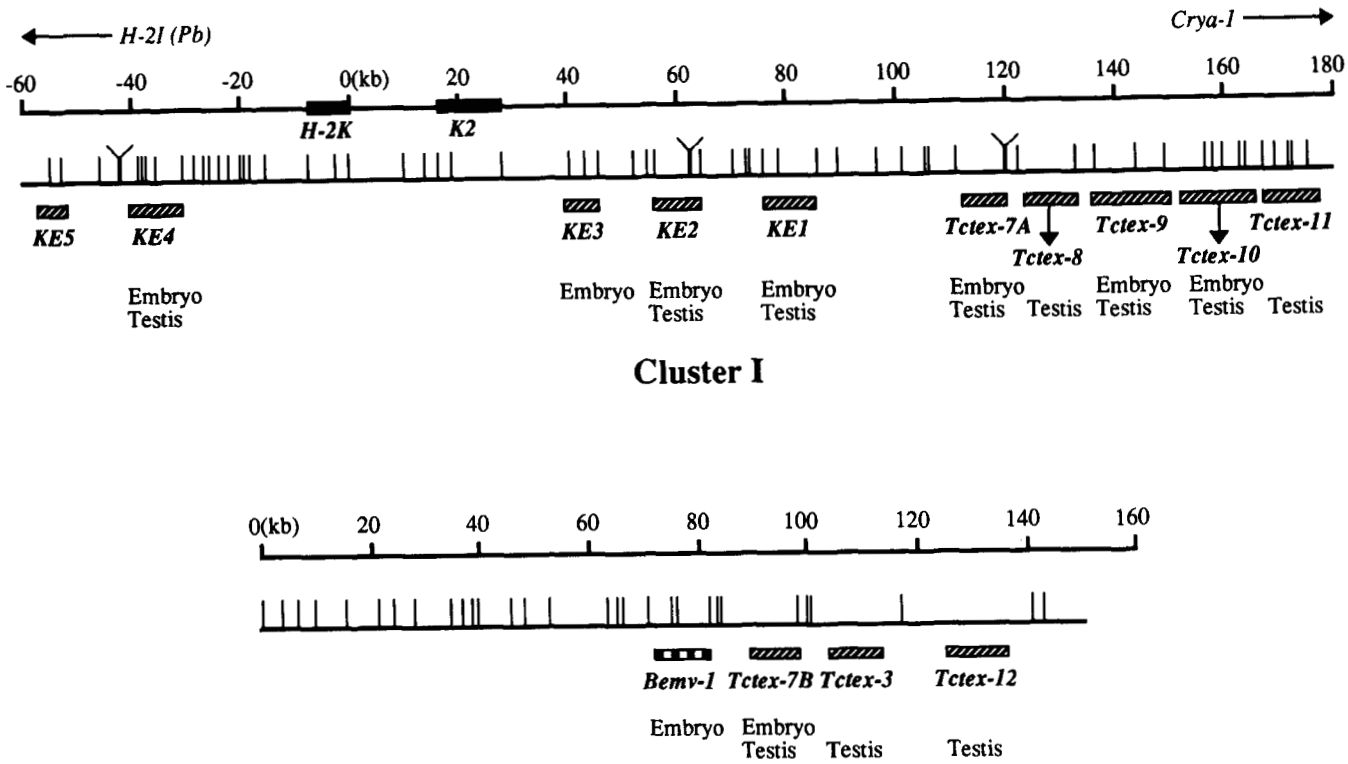


FIGURE 6.—Distribution of transcribed sequences in clusters I and II. The maps are drawn in the same orientation as Figure 2. The approximate position of each gene is indicated by a hatched bar under the *Bam*HI restriction map of the t^{w5} haplotype. The positions of the *H-2K* and *K2* genes are shown as closed bars on the scale indicator. The expression of each gene in embryonic cells and testis is indicated.

the idea that the mutation rate in the H-2K region is very high, and the speculative but intriguing possibility that the high mutation rate is detectable in present-day chromosomes because the H-2K region has a high content of genes transcribed in germ cells and the early embryo.

Sequence change appears intimately associated with genetic activity: Our results suggest the existence of some factors peculiar to the H-2K region which cause high genetic variability. One of the most important factors could be the content of expressed genes in cluster I. It was previously reported that 170 kb out of the 240 kb of cluster I DNA contains four genes expressed during early embryogenesis and one expressed in lymphoid tissues (named *KE* genes) (ABE *et al.* 1988). At least five more genes were found in the newly walked 70 kb of cluster I (YEOM *et al.* 1992). Thus, taken together with the *H-2K* and *K2* genes, the number of transcribed sequences in cluster I totals at least 12 (Figure 6) which averages one every 20 kb. In contrast, the 150 kb of DNA cloned from cluster II encodes only three to four genes transcribed at a level detectable in Northern analysis. (Genomic probes from *Bemv-1* detect a message in Northern analyses, but it is difficult to determine if it comes from this copy of the retrovirus.) There are portions of cluster II which are locally almost as gene dense as cluster I; however, the overall gene density may not be high enough to trigger high sequence variation.

Ample evidence indicates that transcription is intimately associated with sequence changeability of a gene. For example, studies with chemical carcinogens or UV irradiation showed that transcriptionally active regions are much more readily accessible to mutagens or DNA repair enzymes (BOHR, PHILLIPS and HANAWALT 1987, a review). Transcription also facilitates the regional recombination frequency (VOELKEL-MEIMAN, KEIL and ROEDER 1987; BLACKWELL and ALT 1989; NICOLAS *et al.* 1989; THOMAS and ROTHSTEIN 1991, a review) by changing the DNA conformation (GOTTLIEB and ESPOSITO 1989) or by increasing accessibility to recombination machinery (KUNZ and HAYNES 1981).

Also relevant in this context is the fact that the H-2 complex contains several recombinational hot spots (STEINMETZ, UEMATSU and FISCHER-LINDAHL 1987) including one at the H-2I-K boundary region in *t*-haplotypes (−45 to −95 on the same scale as in Figure 6) (ARTZT *et al.* 1988). Thus, since cluster I is densely populated with transcribed sequences, and adjacent to the recombinationally active region, it may well be that the chromatin structure in cluster I is exposed to a higher level of transcription, recombination, and mutation than in cluster II. In support of this UEHARA

et al. (1987) have shown that the leftmost 140 kb of cluster I contains numerous RFLPs between two *t*-haplotypes due to insertions/deletions of murine Alu-like mobile elements.

A chromosomal region with a comparable, high transcriptional activity is found in the human MHC class III region (mouse H-2S region); at least 19 genes have been reported to reside in 600 kb of the class III region (SARGENT, DUNHAM and CAMPBELL 1989; SPIES, BRESNAHAN and STROMINGER 1989). Some of these genes such as *tumor necrosis factor* (*TNF*) α and β, *heat shock protein 70* (*HSP70*), *steroid 21-hydroxylase* (*CYP21*) and complement components *C2*, *C4* and *factor B* have been well characterized. Interestingly, *CYP21*, *C2*, *C4* and *factor B* show extensive polymorphisms in humans (SCHNEIDER *et al.* 1986; CAMPBELL 1987). In addition, *TNF*, *HSP*, *CYP21* and *C4* are present as duplicated sequences within the class III region. Taken together, this might indicate that the class III region is also very prone to mutation.

Expression pattern of genes may dictate whether mutations are germinal or somatic: Although a given region of chromosome with active transcription might have an equal chance of suffering a selectively neutral mutation, the heritability of the mutation would be quite different depending on the genetic environment where it occurs. If a mutation happens in a region dense with genes expressed in germ cells or during early embryogenesis before the germ cells are set aside, it may be more likely to be inherited. In contrast, if the change occurs in a region where the resident genes are scarce or expressed only in somatic cells, it will not be propagated. Most of the *KE* and *Tctex* genes of cluster I are expressed in embryos and/or testis (Figure 6, ABE *et al.* 1988; YEOM *et al.* 1992). Thus, the high polymorphism in cluster I could be the result of an increased germinal mutation rate caused by a synergistic action of two factors: the high gene density, and their expression in present and/or future germ cells. On an evolutionary time scale, these circumstances would give rise to a detectable mutational hot spot.

The implications of a high mutation rate in cluster I for the evolution of the *H-2K* gene: The class I and II MHC molecules are cell surface glycoproteins that function in immune recognition and regulation (KLEIN 1986). In most species studied, these molecules show a high degree of polymorphism in the population. For example, it has been estimated that 50–100 alleles exist for each of the human class I loci, *HLA-A* and *HLA-B*, and over 100 alleles for each of the corresponding mouse loci *H-2K* and *H-2D* (KLEIN and FIGUEROA 1981; KLEIN 1987). The MHC polymorphism was initially explained by assuming that the MHC loci undergo a high rate of mutation (KLEIN 1978). However, later studies based on molecular genetic data presented compelling evidence against the high mutation rate hypothesis. The data indicate that the MHC genes are old, and that in the regions other than the antigen recognition site (ARS), the MHC genes do not change faster than any other nuclear genes (HAYASHIDA and MIYATA 1983; KLEIN 1987). Historically, many hypotheses have been proposed to explain the high degree of the MHC polymorphism. They include *trans*-species hypothesis (KLEIN 1987), overdominant selection (DOHERTY and ZINKERNAGEL 1975; HUGHES and NEI 1988) or positive heterozygosity selection (FLAHERTY 1988), gene conversion by recombination (LOPEZ DE CASTRO *et al.* 1982; OHTA 1982; HAYASHIDA and MIYATA 1983; GELIEBTER and NATHENSON 1987), molecular drive (DOVER 1986), and frequency-dependent selection (SNELL 1968; BODMER 1972).

The majority of the MHC polymorphism comes from the variability in the ARS of the molecule. One can explain the high MHC polymorphism as the result of one or more of the following mechanisms: (1) positive (overdominant) selection for heterozygosity in ARS, (2) relaxed negative selection in ARS, and (3) intrinsically high mutation rate in the MHC loci. First, theoretical considerations predict that if positive selection plays an active role, the mutation rate does not have to be particularly high to produce a high polymorphism. Indeed, the contribution of overdominant selection to the MHC polymorphism has been clearly documented by several authors (FLAHERTY 1988; HUGHES and NEI 1988, 1989). According to HUGHES and NEI (1988, 1989), the rate of nonsynonymous (amino acid-altering) changes of the ARS sequences far exceeds the rate of synonymous changes for the same area of the classical class I and II molecules in both human and mouse. However, the pattern is reversed in the non-ARS of the molecules.

Second, if the selection pressure is less intense than normal in the ARS, the net result may mimic the effects of the positive selection. One expectation is that with the relaxed negative selection pressure acting on the ARS, the rate of synonymous changes would be more or less similar to the rate of nonsynonymous changes. However, this is clearly not the case, and thus this mechanism may not play an important role.

Third, an intrinsically elevated mutation rate will obviously increase the variability by enlarging the pool of variation for selection. There are two factors influencing the level of DNA variation: the mutation rate and natural selection. We are assuming little or no consequence to the fitness effects for the mutations described in this study, and therefore, differences in RFLP frequencies reflect differences in the mutation rate. Considering the random distribution of RFLPs and the location of the *H-2K* gene in cluster I, it is clear that the *H-2K* gene lies at the center of a highly mutable region. Our results suggest that, whatever

phenomena may be involved in the maintenance of MHC polymorphism at the population level, the fundamental mechanism responsible for the high polymorphism of the region surrounding *H-2K* is probably the intrinsically increased rate of germ-line mutations. This view is in apparent contradiction to the fact that, except for the ARS region, the *H-2K* gene has evolved at a rate much like most other genes. It is likely that the *H-2K* gene undergoes a high incidence of mutation at the level of the individual; however, these mutations might be efficiently removed from the population through mechanisms such as selective pressure. Therefore, the non-ARS regions of the molecule are expected to undergo intense selection against mutation because any significant changes in them could hamper the formation of proper three-dimensional conformation and binding with accessory molecules. However, as evidenced by the overdominant selection, the ARS seems to be under a different kind of selection pressure than other regions of the molecule due to the necessity to interact with a diversity of antigens. In fact, as emphasized by FLAHERTY (1988), it must be under a positive selective advantage for heterozygosity, a situation which would favor the survival of mutations leading to the generation of new alleles. The slow evolution rate of the non-ARS of the MHC molecules may be explained by purifying mechanisms.

A caveat that might be added is that the high germline mutation rate in cluster I could be unique to rodents, since in these animals a short DNA stretch encompassing the *H-2K* or equivalent (about 60 kb in the mouse) was translocated (HANSON and TROWSDALE 1991) from the primordial class I gene cluster to a region where testis-/embryo-expressed genes are densely populated (YEOM *et al.* 1992).

There is no information about the polymorphic status of the DNA region flanking the *H-2D* genes. However, given the highly polymorphic nature of the H-2K region, it is interesting to speculate that the *H-2K* gene may be the ultimate generator of the class I molecule diversity which in turn is propagated to other class I loci through genetic mechanisms such as microrecombination (GELIEBTER and NATHENSON 1987). This idea is supported by previous study on the *H-2* haplotype variation in 33 *t*-chromosomes collected from around the world (NIZETIC, FIGUEROA and KLEIN 1984). The study indicates that the H-2K molecule is more diversified than the H-2D molecule in *t*-haplotypes. Moreover, the *K* locus has more minor alleles (alleles with only a slight variation from a major allele) than the *D* locus does (11 *vs.* 3). A similar study on inbred strains also reveals more minor alleles for *H-2K* than for *H-2D* (22 *vs.* 5) (KLEIN and FIGUEROA 1981).

Based on the data presented here, we suggest that the high polymorphism of *H-2K* probably has its root in a high germ-line mutation rate aided by other mechanisms. That genetic environment could be important in the evolution of DNA sequences does not detract from the *trans*-species hypothesis, but rather adds back the element of a high mutation rate.

We thank LORI FLAHERTY for helpful discussions and JIM BULL for critical comments on the manuscript. This work was supported by grants from the National Institutes of Health (CA21651 and HD10668).

LITERATURE CITED

ABE, K., J.-F. WEI, F.-S. WEI, Y.-C. HSU, H. UEHARA, K. ARTZT and D. BENNETT, 1988 Searching for coding sequences in the mammalian genome: the H-2K region of the mouse MHC is replete with genes expressed in embryos. EMBO J. **7:** 3441–3449.

ARTZT, K., P. MCCORMICK and D. BENNETT, 1982 Gene mapping within the *T/t* complex of the mouse. II. Anomalous position of the *H-2* complex in *t*-haplotypes. Cell **28:** 463–470.

ARTZT, K., H.-S. SHIN, D. BENNETT and A. DIMEO-TALENTO, 1985 Analysis of major histocompatibility complex haplotypes of *t*-chromosomes reveals that the majority of diversity is generated by recombination. J. Exp. Med. **162:** 93–104.

ARTZT, K., K. ABE, H. UEHARA and D. BENNETT, 1988 Intra-*H-2* recombination in *t*-haplotypes shows a hot spot and close linkage of 1^{tw5} to *H-2K*. Immunogenetics **28:** 30–37.

BENNETT, D., 1975 The *T*-locus of the mouse: a review. Cell **6:** 441–454.

BLACKWELL, T. K., and F. W. ALT, 1989 Mechanism and developmental program of immunoglobulin gene rearrangement in mammals. Annu. Rev. Genet. **23:** 605–636.

BODMER, W. F., 1972 Evolutionary significance of the *HL-A* system. Nature **237:** 139–145.

BOHR, V. A., D. H. PHILLIPS and P. C. HANAWALT, 1987 Heterogeneous DNA damage and repair in the mammalian genome. Cancer Res. **47:** 6426–6436.

CAMPBELL, R. D., 1987 The molecular genetics and polymorphism of *C2* and *factor B*. Br. Med. Bull. **43:** 37–49.

DOHERTY, P. C., and R. M. ZINKERNAGEL, 1975 Enhanced immunological surveillance in mice heterozygous at the *H-2* gene complex. Nature **256:** 50–52.

DOVER, G. A., 1986 Molecular drive in multigene families: how biological novelties arise, spread, and are assimilated. Trends Genet. **2:** 159–165.

FEINBERG, A. P., and B. VOGELSTEIN, 1983 A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:** 6–13.

FIGUEROA, F., M. GOLUBIC, D. NIZETIC and J. KLEIN, 1985 Evolution of mouse major histocompatibility complex genes borne by *t*-chromosomes. Proc. Natl. Acad. Sci. USA **82:** 2819–2823.

FLAHERTY, L., 1988 Major histocompatibility complex polymorphism: a nonimmune theory for selection. Hum. Immunol. **21:** 3–13.

GELIEBTER, J., and S. G. NATHENSON, 1987 Recombination and the concerted evolution of the murine MHC. Trends Genet. **3:** 107–112.

GOLUBIC, M., F. FIGUEROA, M. TOSI and J. KLEIN, 1984 Restriction fragment length polymorphism of *C4* genes in mice with *t* chromosomes. Immunogenetics **21:** 247–256.

GOTTLIEB, S., and R. E. ESPOSITO, 1989 A new role for a yeast transcriptional silencer gene, *SIR2*: in regulation of recombination in ribosomal RNA. Cell **56:** 771–776.

HA, H.-S., C. A. HOWARD, Y. I. YEOM, K. ABE, H. UEHARA, K. ARTZT and D. BENNETT, 1991 Several testis-expressed genes in the mouse *t*-complex have expression differences between wild-type and *t*-mutant mice. Dev. Genet. **12:** 318–332.

HAMMER, M. F., J. SCHIMENTI and L. SILVER, 1989 Evolution of mouse chromosome *17* and the origin of inversions associated with *t*-haplotypes. Proc. Natl. Acad. Sci. USA **86:** 3261–3265.

HANSON, I. M., and J. TROWSDALE, 1991 Colinearity of novel genes in the class II regions of the MHC in mouse and human. Immunogenetics **34:** 5–11.

HAYASHIDA, H., and T. MIYATA, 1983 Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. Proc. Natl. Acad. Sci. USA **80:** 2671–2675.

HERRMANN, B., M. BUCAN, P. E. MAINS, A.-M. FRISCHAUF, L. M. SILVER and H. LEHRACH, 1986 Genetic analysis of the proximal portion of the mouse *t*-complex: evidence for a second inversion within *t*-haplotypes. Cell **44:** 469–476.

HUGHES, A. L., and M. NEI, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335:** 167–170.

HUGHES, A. L., and M. NEI, 1989 Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA. **86:** 958–962.

KLEIN, J., 1978 *H-2* mutations: their genetics and effect on immune functions. J. Adv. Immun. **26:** 55–146.

KLEIN, J., 1986 *Natural History of the Major Histocompatibility Complex*. John Wiley & Sons, New York.

KLEIN, J., 1987 Origin of major histocompatibility complex polymorphism: the *trans*-species hypothesis. Hum. Immunol. **19:** 155–162.

KLEIN, J., and F. FIGUEROA, 1981 Polymorphism of the mouse *H-2* loci. Immunol. Rev. **60:** 23–57.

KLEIN, J., P. SIPOS and F. FIGUEROA, 1984 Polymorphism of *t*-complex genes in European wild mice. Genet. Res. **44:** 39–46.

KUNZ, B. A., and R. H. HAYNES, 1981 Phenomenology and genetic control of mitotic recombination in yeast. Annu. Rev. Genet. **15:** 57–89.

LOPEZ DE CASTRO, J. A., J. L. STROMINGER, D. M. STRONG and H. T. ORR, 1982 Structure of crossreactive human histocompatibility antigens *HLA-A28* and *HLA-A2*: possible implications for the generation of *HLA* polymorphism. Proc. Natl. Acad. Sci. USA **79:** 3813–3817.

MANN, E. A., L. M. SILVER and R. W. ELLIOTT, 1986 Genetic analysis of a mouse *T* complex locus that is homologous to a kidney cDNA clone. Genetics **114:** 993–1006.

NICOLAS, A., D. TRECO, N. P. SCHULTES and J. W. SZOSTAK, 1989 An initiation site for meiotic gene conversion in the yeast *Saccharomyces cerevisiae*. Nature **338:** 35–39.

NIZETIC, D., F. FIGUEROA and J. KLEIN, 1984 Evolutionary relationships between the *t* and *H-2* haplotypes in the house mouse. Immunogenetics **19:** 311–320.

OHTA, T., 1982 Allelic and nonallelic homology of a supergene family. Proc. Natl. Acad. Sci. USA **79:** 3251–3254.

RACKWITZ, H. R., G. ZEHETNER, A.-M. FRISCHAUF and H. LEHRACH, 1984 Rapid restriction mapping of DNA cloned in lambda phage vectors. Gene **30:** 195–200.

RACKWITZ, H. R., G. ZEHETNER, H. MURIALDO, H. DELIUS, J. H. CHAI, A. POUSTKA, A.-M. FRISCHAUF and H. LEHRACH, 1985 Analysis of cosmids using linearization by phage lambda terminase. Gene **40:** 259–266.

RIGBY, P. W. J., M. DIECKMANN, C. RHODES and P. BERG, 1977 Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113:** 237–251.

ROEHME, D., H. FOX, B. HERRMANN, A.-M. FRISCHAUF, J.-E. EDSTROM, P. MAINS, L. M. SILVER and H. LEHRACH, 1984 Molecular clones of the mouse *t*-complex derived from microdissected metaphase chromosomes. Cell **36:** 783–788.

SAGE, R. D., 1981 Wild mice, pp. 39–90 in *The Mouse in Biomedical Research*, Vol. 1, edited by H. L. FOSTER, J. D. SMALL and J. G. FOX. Academic Press, New York.

SARGENT, C. A., I. DUNHAM and R. D. CAMPBELL, 1989 Identification of multiple HTF-island associated genes in the human major histocompatibility complex class III region. EMBO J. **8:** 2305–2312.

SCHNEIDER, P. M., M. C. CARROLL, C. A. ALPER, C. RITTNER, A. S. WHITEHEAD, E. J. YUNIS and H. R. COLTEN, 1986 Polymorphism of the human complement *C4* and *steroid 21-hydroxylase* genes. J. Clin. Invest. **78:** 650–657.

SHIN, H.-S., 1989 The *T/t* complex and the genetic control of mouse development, pp. 443- 471 in *Human Immunogenetics: Basic Principles and Clinical Relevance*, edited by S. D. LITWIN. Marcel Dekker, New York.

SHIN, H.-S., J. STAVNEZER, K. ARTZT and D. BENNETT, 1982 The genetic structure and origin of *t*-haplotypes of mice analyzed with *H-2* cDNA probes. Cell **29:** 969–976.

SHIN, H.-S., L. FLAHERTY, K. ARTZT, D. BENNETT and J. RAVETCH, 1983 Inversion in the *H-2* complex in *t*-haplotypes of mice. Nature **306:** 380–383.

SILVER, L. M., 1982 Genomic analysis of the *H-2* complex region associated with mouse *t*-haplotypes. Cell **29:** 961–968.

SILVER, L. M., 1985 Mouse *t*-haplotypes. Annu. Rev. Genet. **19:** 179–208.

SILVER, L. M., and K. ARTZT, 1981 Recombination suppression of mouse *t*-haplotypes is due to chromatin mismatching. Nature **290:** 68–70.

SILVER, L. M., M. HAMMER, H. FOX, J. GARRELS, M. BUCAN, B. HERRMANN, A.-M. FRISCHAUF, H. LEHRACH, H. WINKING, F. FIGUEROA and J. KLEIN, 1987 Molecular evidence for the rapid propagation of mouse *t*-haplotypes from a single, recent, ancestral chromosome. Mol. Biol. Evol. **4:** 473–482.

SILVER, L. M., J. UMAN, J. DANSKA and J. I. GARRELS, 1983 A diversified set of testicular cell proteins specified by genes within the mouse *t*-complex. Cell **35:** 35–45.

SNELL, G. D., 1968 The *H-2* locus of the mouse: observations and speculations concerning its comparative genetics and its polymorphism. Folia Biol. **14:** 335–358.

SPIES, T., M. BRESNAHAN and J. L. STROMINGER, 1989 Human major histocompatibility complex contains a minimum of 19 genes between the complement cluster and *HLA-B*. Proc. Natl. Acad. Sci. USA **86:** 8955–8958.

STEINMETZ, M., Y. UEMATSU and K. FISCHER-LINDAHL, 1987 Hotspots of homologous recombination in mammalian genomes. Trends Genet. **3:** 7–10.

STEINMETZ, M., M. MALISSEN, L. HOOD, A. ORN, R. A. MAKI, G. R. DASTOORNIKOO, D. STEPHAN, E. GIBB and R. ROMANIUK, 1984 Tracts of high or low sequence divergence in the mouse major histocompatibility complex. EMBO J. **3:** 2995–3003.

STURM, S., F. FIGUEROA and J. KLEIN, 1982 The relationship between *t* and *H-2* complexes in wild mice. I. The *H-2* haplotypes of 20 *t*-bearing strains. Genet. Res. **40:** 73–88.

THOMAS, B. J., and R. ROTHSTEIN, 1991 Sex, maps and imprinting. Cell **64:** 1–3.

UEHARA, H., K. ABE, C.-H. T. PARK, H.-S. SHIN, D. BENNETT and K. ARTZT, 1987 The molecular organization of the H-2K region of two *t*-haplotypes: implications for the evolution of genetic diversity. EMBO J. **6:** 83–90.

VOELKEL-MEIMAN, K., R. L. KEIL and G. S. ROEDER, 1987 Recombination-stimulating sequences in yeast ribosomal DNA correspond to sequences regulating transcription by RNA polymerase I. Cell **48:** 1071–1079.

YEOM, Y. I., H.-S. HA, R. BALLING, H. R. SCHÖLER and K. ARTZT, 1991 Structure, expression and chromosomal location of the *Oct-4* gene. Mech. Dev. **35:** 171–179.

YEOM, Y. I., K. ABE, D. BENNETT and K. ARTZT, 1992 Testis-/embryo-expressed genes are clustered in the mouse H-2K region. Proc. Natl. Acad. Sci. USA (in press).

Communicating editor: W.-H. LI